

M140

Written EMA

2019B

Covers the whole module

Cut-off date 11 September 2019

The end-of-module assessment (EMA) consists of **two parts**: this written EMA and the iCME (called ‘iCME 81’).

Before you begin work on this written EMA, please read the instructions on the module website for preparing and submitting the EMA. Note in particular:

- you should send this part of the EMA to the address provided by The Open University (*not your tutor*) or submit it electronically using the University’s online TMA/EMA service
- your tutor is not permitted to grant an extension for your EMA.

If you submit the EMA electronically, then you must submit it as a **PDF** file.

This part of the EMA is marked out of 70. The marks allocated to each part of each question are indicated in brackets in the margin.

The Minitab worksheet that you require for this part of the EMA should be downloaded from the ‘Assessment’ area on the module website.

Unless otherwise specified in the question, it is up to you whether you use Minitab to arrive at your answer.

Remember that this is just one part of the EMA, and you must submit **both parts** of the EMA by the respective cut-off dates.

This assessment covers *the whole module*.

The questions in this EMA all consider data associated with orienteering. Orienteering is a sport where participants aim to navigate their way round a series of points in the shortest time possible. The locations of the points are given on a map, so doing well depends on map-reading skill as well as running ability. Orienteering attracts competitors across a wide range of ages, so usually different courses are laid on to ensure that the physical demands are reasonable for all.

In this EMA you will analyse some data from one course of an orienteering competition that was held in 2017. This competition was split across six days, with the places attained by each competitor on each day combined to determine the overall winner.

Question 1 – 3 marks

The Minitab worksheet **orienteering2017.mtw** contains the following variables:

- **id**: the serial number that identifies each individual participant ('orienteer') in the orienteering competition.
 - **ageclass**: the gender/age class for each participant, categorised as 'M70' for men aged 70–74, 'W55' for women aged 55–59, 'W60' for women aged 60–64, and 'W70' for women aged 70–74.
 - **time1**: the time (in minutes) that the orienteer took to complete the course on day 1 of the competition
 - **time2**: the time (in minutes) that the orienteer took to complete the course on day 2 of the competition
- (a) Which orienteer was the fastest on day 1? That is, which orienteer had the lowest time on day 1? To which gender/age class did this orienteer belong? [2]
- (b) Was the fastest orienteer on day 1, who you identified in part (a), also the fastest of all the orienteers on day 2? Was this participant the fastest orienteer in their gender/age class on day 2? [1]

Question 2 – 5 marks

In this question, you will use the data given in the Minitab worksheet **orienteering2017.mtw** to investigate the distribution of times on day 1 of the competition.

- (a) Using Minitab, produce a stemplot of the times on day 1. Include this stemplot in your answer. (In doing this, you should leave the **Increment** field blank and the **Trim outliers** option unselected.) [1]
- (b) Use your stemplot to describe the shape of the times recorded on day 1. Justify your answer. [4]

Question 3 – 18 marks

The times taken by competitors on the course on day 1 for the different gender/age classes (M70, W55, W60 and W70) are given in separate columns in the Minitab worksheet **timesbyclass.mtw**.

- (a) Using Minitab, or otherwise, produce a diagram containing boxplots of the times taken by competitors on this course on day 1 for the different gender/age classes (M70, W55, W60 and W70). You should ensure that the boxplots are horizontal and drawn on the same scale. You should prepare the boxplots ready for inclusion in a report by ensuring that the title and horizontal-axis label are clear and informative. Include the finished boxplots in your answer. [5]
- (b) Complete the following table of summary statistics for times taken on day 1 for each of the gender/age classes. (You should give the numbers of orienteers to the nearest whole number, and all the other values rounded to two decimal places.) [2]

	M70	W55	W60	W70
Number				
Mean				
Median				
Standard deviation				
Interquartile range				
Range				

- (c) Using your answers to both parts (a) and (b), does the time taken on the course on day 1 appear to be the same, on average, in the different gender/age classes? Justify your opinion. [3]
- (d) Suppose that it could be argued that as the gender/age classes M70 and W70 are on the same course, they will have the same population mean time to complete the course. Write down suitable null and alternative hypotheses to test this claim, stating clearly the meanings of any symbols that you use. [2]
- (e) A two-sample t -test (for populations with a common variance) could be used to test the hypotheses that you wrote down in part (d). The appropriateness of this test depends on two assumptions. State these two assumptions. Are these assumptions reasonable in this case? Use your answers to parts (a) and (b) to justify your opinion. [6]

Question 4 – 12 marks

In the Minitab worksheet **orienteering2017.mtw**, the times taken by the orienteers on day 2 are given as well as the times taken on day 1.

- (a) Obtain a scatterplot of the times taken on day 2 against the times taken on day 1, putting the times taken on day 2 on the vertical axis, and the times taken on day 1 on the horizontal axis. Include this scatterplot in your answer. Interpret this scatterplot, giving reasons for your interpretation. [9]
- (b) Obtain the correlation coefficient between the times on day 1 and the times on day 2. With reference to the scatterplot that you produced in part (a), state whether this correlation coefficient provides a good representation of the strength of the relationship between the actual times and those predicted by the model. Justify your answer. [3]

Question 5 – 9 marks

It was intended that the courses on days 1 and 2 of the competition would be of the same difficulty. One way of judging this is to see whether the average difference in times for runners on the two days is zero.

In this question you will use a one-sample z -test to investigate whether, based on the data given in the Minitab worksheet **orienteering2017.mtw**, days 1 and 2 had similar difficulty according to this criterion. For these data, the differences in times taken by a competitor between day 1 and day 2 (using `time2 – time1`) have mean 8.14 and standard deviation 17.42.

Throughout this question you should do all calculations by hand and show all your working.

- (a) Write down suitable null and alternative hypotheses, stating clearly the meanings of any symbols that you use. [2]
- (b) Calculate the value of the estimated standard error of the mean difference in times taken between day 1 and day 2 (using `time2 – time1`). [2]
- (c) Calculate the value of the test statistic for the one-sample z -test. [2]
- (d) Complete the hypothesis test, carefully detailing the conclusions of the test. [3]

Question 6 – 13 marks

The data given in the Minitab worksheet **orienteering2017.mtw** correspond only to orienteers who completed all six days of the competition. There were others who failed to complete one or more of these days (either because they failed to complete the course correctly or because they did not compete on those days). The following table shows the number of orienteers in each gender/age class who completed the course on 0, 1, ... or 6 days.

Number of days completed	M70	W55	W60	W70	Total
6	16	28	17	21	82
5	4	5	7	11	27
4	6	5	3	4	18
3	2	3	4	2	11
2	1	2	2	0	5
1	3	0	1	0	4
0	1	1	0	0	2
Total	33	44	34	38	149

- (a) Explain why this table is correctly described as a contingency table. [3]
- (b) What is the probability that a randomly selected orienteer was in the W55 gender/age class and completed exactly five days? Give your answer to three decimal places. [1]
- (c) What is the probability that a randomly selected orienteer completed all six days if it is known that the orienteer was in the W70 gender/age class? Give your answer to three decimal places. [1]
- (d) Suppose that a researcher is interested in using this sample to investigate whether there is any relationship or association between the number of days completed and the gender/age class. To do this, the researcher will carry out a χ^2 test using a contingency table in which the 'Number of days completed' categories '0', '1', '2', '3' and '4' are combined into the single category '4 or fewer'. The resulting combined contingency table is shown below.

Number of days completed	M70	W55	W60	W70	Total
6	16	28	17	21	82
5	4	5	7	11	27
4 or fewer	13	11	10	6	40
Total	33	44	34	38	149

The data for this contingency table are given in the Minitab worksheet **completion.mtw**.

- (i) Write down suitable null and alternative hypotheses for this χ^2 test. [2]
- (ii) Use Minitab to carry out the χ^2 test for contingency tables to test the hypotheses that you wrote down in part (d)(i). Include a copy of the Minitab output for the test in your answer. [2]
- (iii) Explain why it was not valid to apply the χ^2 test to the full contingency table, but it is valid to apply the χ^2 test to the combined contingency table given in the Minitab worksheet **completion.mtw**. [2]
- (iv) What do you conclude from the χ^2 test? [2]

Question 7 – 10 marks

A group of sports psychologists is interested in finding out whether a new technique that they've developed will help older orienteers to maintain a positive mental attitude while competing, and hence will improve their performance in the long term.

The sports psychologists decide to run a trial in which they teach their technique to a sample of older orienteers. They intend to select a group of orienteers to whom they will teach the technique, and will also select another group that they will just monitor.

- (a) What design of clinical trial best describes the study that the sports psychologists intend to set up? [1]
 - (b) One of the variables that the sports psychologists will measure is the time taken for each orienteer to go round a particular orienteering course.
 - (i) Give a reason why this variable could be regarded as objective data. [1]
 - (ii) Give a reason why this variable would be regarded as interval scale data. [1]
 - (c) Suggest a hypothesis test that could be used to analyse the data described in part (b) to investigate whether there is evidence that the psychological test improves performance. Give a reason why this test might turn out to be unsuitable. [2]
 - (d) The sports psychologists are also interested in the attitude of older orienteers towards sports psychology. They intend to assess this by conducting in-depth interviews with some of the orienteers represented in the Minitab worksheet **orienteering2017.mtw**.
 - (i) Suppose that the sports psychologists have the time and resources to interview roughly only one-twelfth of the competitors. Select, by hand, a suitable sample for the sports psychologists, using systematic random sampling. Any random numbers that you require should be obtained from the random number table given in Unit 4, starting at the beginning of row **68**. Show your working. [3]
 - (ii) In the Minitab worksheet **orienteering2017.mtw**, the orienteers are grouped (vertically) by gender/age class. Explain why the sample that you obtained in part (d)(i) is similar to a stratified sample. In what way is it different? [2]
-